sageza

Strategic Snapshot

File System Archiving:
Because Not All Data Is Equal in Value

By Joyce Tompsett Becknell

The Sageza Group, Inc.
June 2006

The Sageza Group, Inc.
32108 Alvarado Blvd #354
Union City, CA 94587
510·675·0700 fax 650·649·2302
London +44 (0) 20·7900·2819
Milan +39 02·9544·1646

sageza.com
info@sageza.com

File System Archiving:
Because Not All Data Is Equal in Value

ABSTRACT

Data management has historically been the sole responsibility of information technology (IT) managers, but the importance of data to the success and survival of the business is now capturing the interest of business managers and executives as well. On the IT side, there are several issues that compel managers to examine their storage strategies. These include the need to lower overall IT costs, which can be affected by improving storage utilization; decreasing backup and recovery times; and simplifying management tasks to drive higher productivity of IT staff. Seeking answers for these issues has been the impetus behind the adoption of Information Lifecycle Management (ILM) strategies within enterprise and mid-range organizations, to make sure that the storage infrastructure is aligned to the relative importance of the data at any given point in its lifecycle.

On the business side, several factors are driving the growing interest in the state of corporate data. Employee productivity is a concern as today's knowledge workers generate enormous amounts of data critical to the functioning of the business in the form of spreadsheets, presentations, documents, and html files, as well as audio and video files. This data must be managed appropriately so that workers can retrieve and use it when necessary, while keeping storage and storage management costs in check. In addition, the growth in regulatory and compliance requirements means that certain data must be retained for specific periods of time. That data may not be accessed often if at all; but it needs to be digitally maintained, sometimes for decades. Finally, litigation may require a company to produce data as evidence. Data must be accessible when these situations arise, though it may not have been accessed for a long time.

File System Archiving:
Because Not All Data Is Equal in Value

TABLE OF CONTENTS

The Growing Importance of Data Management

Data management has historically been the sole responsibility of information technology (IT) managers, but the importance of data to the success and survival of the business is now capturing the interest of business managers and executives as well. On the IT side, there are several issues that compel managers to examine their storage strategies. These include the need to lower overall IT costs, which can be affected by improving storage utilization; decreasing backup and recovery times; and simplifying management tasks to drive higher productivity of IT staff. Seeking answers for these issues has been the impetus behind the adoption of Information Lifecycle Management (ILM) strategies within enterprise and mid-range organizations, to make sure that the storage infrastructure is aligned to the relative importance of the data at any given point in its lifecycle.

On the business side, several factors are driving the growing interest in the state of corporate data. Employee productivity is a concern as today's knowledge workers generate enormous amounts of data critical to the functioning of the business in the form of spreadsheets, presentations, documents, and html files, as well as audio and video files. This data must be managed appropriately so that workers can retrieve and use it when necessary, while keeping storage and storage management costs in check. In addition, the growth in regulatory and compliance requirements means that certain data must be retained for specific periods of time. That data may not be accessed often if at all; but it needs to be digitally maintained, sometimes for decades. Finally, litigation may require a company to produce data as evidence. Data must be accessible when these situations arise, though it may not have been accessed for a long time.

Because not all data is equal in value, and because the importance of data changes over the course of its lifetime, many IT organizations are considering tiered storage as a way to balance data growth and lower costs, while providing the higher levels of service required in today's business environment. The idea behind tiered storage—and ILM in general—is that only a company's newest or most critical data is reserved on the highest-performing and therefore more expensive storage systems. Aged data or data that is used less frequently can be moved—or archived—to less expensive disk systems, while maintaining quick retrieval times. If retention and security requirements for data are mandated by government regulations, data may be archived to a specialized device that guarantees compliance. Other options include migrating data to tape or optical storage for long-term or "deep" archiving. While this is a sensible approach, the challenge for storage managers is to define policies to properly group data and move it to the correct storage device. Automating this process is optimal, and there are technologies available to facilitate this approach.

File System Archiving (FSA) solutions do provide automated policy-based data movement. Organizations use FSA technologies to move data to lower-cost storage when its access characteristics change to inactive or infrequent. FSA solutions also allow managers to move data to an archive for business governance or legal compliance reasons and to restore data directly to the end-user or application when requested. In this paper, we take a look at one particular FSA solution: EMC's DiskXtender file system archiving software. This product is being deployed in a range of hardware configurations, and we explore those possibilities. We also take a look at two examples of customer deployments of DiskXtender, one in health care and one in financial services, to get more concrete examples of how file system archiving can help businesses make better use of their infrastructure, increase staff productivity, and rest assured that their files will be accessible when needed.

A Closer Look at EMC's DiskXtender Offerings

Organizations cite two reasons in particular to archive files from production capacity to other tiers of storage: files are currently inactive but may be needed for future reference, or there may be a legal or regulatory reason for maintaining the files for a set period of time. EMC provides file system archiving with its DiskXtender family of software products. EMC DiskXtender automates file system archive management through policies defined by the user. Ideally, policy-based management helps IT departments better meet service levels and maintain the balance between storage costs, capacity utilization, and access requirements.

With EMC's solution, administrators define rules to identify files for migration according to file size, age, last access date, or other file attribute criteria. EMC DiskXtender automatically moves files to secondary tiers of storage, and purges files or deletes files when defined criteria are met. To initiate this process, EMC DiskXtender scans disk systems based upon the policies established, looking for files that meet policy-based attributes. Files are moved to archive locations transparently to the user. When a file is needed, it can either be retrieved from archive storage or read directly from the archive, which saves on primary storage usage. The name "DiskXtender" represents this capability, since the product virtually extends primary disk volumes to secondary storage devices, while users keep the same view of migrated files.

Policies will also contain information about where files should be moved, and different file types may have different policies. Archive storage can be SCSI, Fibre Channel, or ATA disk that is attached directly to a server, to a storage area network (SAN), or to a network-attached storage (NAS) device. Files are likely to be moved to a cost-effective tier of ATA disk, to a specialized device such as EMC Centera content-addressed storage for compliance, or even tape or optical storage. With EMC DiskXtender, up to four storage devices can be written to, based on a single policy. This adds layers of data protection for archived files, which may be required by business best practices or government regulations. By continually moving reference or compliance data to secondary storage devices for archiving, EMC DiskXtender provides virtually unlimited primary storage capacity.

EMC DiskXtender in Complementary Solutions

In conjunction with EMC Symmetrix and CLARiiON disk offerings, DiskXtender can be an enabler of tiered storage, moving files between SCSI- and Fibre Channel-based primary disk systems to more cost-effective ATA platforms. In these configurations, DiskXtender minimizes application downtime and performance degradation, issues that are commonly associated with high disk space consumption.

EMC DiskXtender software also automates the movement of inactive or fixed-content files from primary storage to the more cost-effective EMC Centera CAS. Centera is a disk-based technology that provides fast, online access for archived data. It incorporates software that uniquely "fingerprints" each piece of content, ensuring authenticity and ease of retrieval. Pairing Centera CAS with DiskXtender software creates a solution to ensure maximum integrity of archived files for their full retention periods.

NAS systems allow customers to consolidate IT environments and simplify management. DiskXtender facilitates automated file system archiving from EMC Celerra NAS that has SCSI or Fibre Channel disk to ATA-based Celerra systems, and from Celerra to Centera or other devices. For NetWorker Customers, DiskXtender is a complementary product that can speed backups and recoveries and should be considered as part of an overall backup, recovery, and archiving strategy.

Keep in mind that end users do not need to know where archived files are, because they view them as sitting in the location they occupied before migration. Therefore users access them exactly the way they were accessed previous to being moved to other storage devices. When files are migrated, DiskXtender places small file stubs that point to the file at its new location.

To understand best how file system archiving can benefit companies, we present two examples of real customers using EMC's DiskXtender in their organizations.

File System Archiving in Real World Scenarios

Two industries with rapidly growing data storage needs are health care and financial services. Both industries must balance data storage requirements with cost conservation, and at the same time, both industries are facing increasing governmental and regulatory requirements as well as customer demands that mean information retention and retrieval requirements must be met to maintain service levels and stay in business. EMC has successfully implemented DiskXtender-based solutions in both industries, to help organizations meet their business needs, and we take a look here at a customer in each industry to understand how they have implemented file system archiving.

Scenario One: Health Care Company

This first example is a research center attached to a university hospital. The research center collects patient body scans on a regular basis and compares them over time, so it generates a large volume of data and must refer back to it periodically. The center also performs functional imaging that captures images of body parts in motion, generating several gigabytes of data over a period of a few minutes.

Previously, the customer used a combination of conventional disk, backup, and manual data movement strategies to manage this data, but started to have problems with scalability at about the one terabyte mark. Backups and restores were growing lengthy, there was increasing risk of data loss due to system crashes, and researchers wanted the ability to access archived files without the help of IT. The customer turned to EMC for a solution that was fast, reliable, and cost-efficient for storage and retrieval. The customer decided on a tiered storage system that would provide the ability to store data, retrieve it, and dynamically manage it while maintaining free disk space on a Linux platform, which was the organizational standard. EMC offered the solution of DiskXtender for Unix/Linux that manages the storage process from a central point by prioritizing and moving data to the most suitable system device. Data that is frequently used remains on the disk cache for immediate access, and inactive information is migrated to a robotic tape system until recall.

The DiskXtender solution took about three months to implement, and is now handling storage that increases by about 750GB per month. The solution also allows each department to be viewed differently, so that the organization may place budget requirements sensibly.

Scenario Two: Financial Services Company

The second example is a financial services company that is a provider of funds-based life insurance. It is one company in a portfolio that offers private composite business. The company found its storage requirements increasing greatly, in particular due to government regulations requiring that all computer-generated, tax-relevant data be kept in electronic form for public auditing for ten years. The company launched a project to meet these requirements, and one of its goals was to find a suitable platform for long-term archiving, as the magneto-optical storage system then in use had reached its capacity limit.

The company was already an EMC customer, and it decided to add Centera content-addressed storage with DiskXtender for Windows to meet availability, scalability, performance, cost-effectiveness, and data security requirements. The archive was audited by an independent body for compliance with the applicable trade and tax regulations.

DiskXtender ensures that financial data is seamlessly moved to Centera. The financial data is transferred from a mainframe to a Windows server. There DiskXtender for Windows serves as a data manager to ensure the files are stored on Centera. Centera Governance Edition uses the rules and policies of DiskXtender for Windows to guarantee compliance with the appropriate storage retention periods. Within these defined periods, Centera blocks the deletion or modification of objects. For disaster recovery purposes, the archived data is automatically replicated from Centera via TCP/IP over the network to a second system in another firewall section.

Only a limited number of employees have access to the Centera system. A new document management system was introduced to allow administrative employees to work with data from the long-term archive.

File System Archiving: An Important Part of ILM

More and more customers are considering ways to implement tiered storage as part of their ILM strategy. Certainly those who are familiar with ILM will understand the benefits of tiered storage, but any IT administrator who is facing growing storage needs, backup windows that are lengthening, cost issues, regulatory or compliance requirements, legal concerns, or user productivity issues should investigate the use of tiered storage to better match data value needs to technology capabilities. IT managers are also implementing archives to store inactive files, maintain records, or manage other data that must be preserved for extended periods of time. EMC's DiskXtender product is a natural fit for these environments, as it provides automated, policy-based file system archiving that matches the data to the storage at the right time in its lifecycle according to the company's business rules.

Because DiskXtender software is compatible with a wide range of disk, tape, and optical technologies, it fits into most environments without the need to rip and replace existing infrastructure. It also works transparently to end users so there is no need to change the way they retrieve data currently. Business and IT executives with burgeoning storage, who need to keep costs balanced with user storage demands, should explore file system archiving, and EMC's solution as an extension of their existing ILM strategy. Organizations facing governmental or industry requirements or who have legal requirements for maintaining files will also find that DiskXtender, particularly when combined with EMC's Centera, is an accepted alternative, and can help them meet requirements and stay compliant. In addition, EMC's solution can help storage managers achieve better response times, shrink backup windows, decrease retrieval times, and increase service levels within their existing architecture.